Behavioral/Cognitive

Elucidating the Neural Representation and the Processing Dynamics of Face Ensembles

Tyler Roberts, Jonathan S. Cant,* and Adrian Nestor*
Department of Psychology at Scarborough, University of Toronto, Toronto, Ontario M1C 1A4, Canada

Extensive behavioral work has documented the ability of the human visual system to extract summary representations from face ensembles (e.g., the average identity of a crowd of faces). Yet, the nature of such representations, their underlying neural mechanisms, and their temporal dynamics await elucidation. Here, we examine summary representations of facial identity in human adults (of both sexes) with the aid of pattern analyses, as applied to EEG data, along with behavioral testing. Our findings confirm the ability of the visual system to form such representations both explicitly and implicitly (i.e., with or without the use of specific instructions). We show that summary representations, rather than individual ensemble constituents, can be decoded from neural signals elicited by ensemble perception, we describe the properties of such representations by appeal to multidimensional face space constructs, and we visualize their content through neural-based image reconstruction. Further, we show that the temporal profile of ensemble processing diverges systematically from that of single faces consistent with a slower, more gradual accumulation of perceptual information. Thus, our findings reveal the representational basis of ensemble processing, its fine-grained visual content, and its neural dynamics.

Key words: face perception; visual ensembles; multivariate analysis; neural decoding

Significance Statement

Humans encounter groups of faces, or ensembles, in a variety of environments. Previous behavioral research has investigated how humans process face ensembles as well as the types of summary representations that can be derived from them, such as average emotion, gender, and identity. However, the neural mechanisms mediating these processes are unclear. Here, we demonstrate that ensemble representations, with different facial identity summaries, can be decoded and even visualized from neural data through multivariate analyses. These results provide, to our knowledge, the first detailed investigation into the status and the visual content of neural ensemble representations of faces. Further, the current findings shed light on the temporal dynamics of face ensembles and its relationship with single-face processing.

Introduction

Humans routinely encounter groups of faces in a variety of settings. Whether it be in an office, on a subway, or walking down a busy street, we are sometimes forced to process quickly a large amount of facial information. Given the limits of working memory (Luck and Vogel, 1997), a mechanism has been suggested that allows individuals to encode information, such as size (Chong and Treisman, 2003, 2005), orientation (Dakin and Watt, 1997), and direction of motion (Watamaniuk et al., 1989) from similar objects into averages or "summary representations." Interestingly, such a mechanism also extends to the processing of more complex properties, including, in the case of faces, emotional expression (Haberman and Whitney, 2007, 2009), gaze (Sweeny and Whitney, 2014; Florey et al., 2016), gender (Haberman and Whitney, 2007), and even identity (de Fockert and Wolfenstein, 2009; Neumann et al., 2013, 2017; Leib et al., 2014; Haberman et al., 2015).

Despite growing interest in the behavioral study of facial summary representations (Whitney and Yamanashi Leib, 2018), the neural underpinnings of ensemble face encoding have received far less attention. Recent fMRI work (Im et al., 2017) has suggested that the perception of individual face emotions and that of crowd emotions recruit differentially the two visual streams, with the former relying more on the ventral stream and the latter relying more on the dorsal stream. However, ventral areas, such as the lateral occipital area and the parahippocampal place area, have also been critically implicated in object ensemble processing (Cant and Xu, 2012, 2017). Regarding the dynamics of ensemble perception, EEG work (Puce et al., 2013) has shown that event-

Received March 1, 2019; revised Aug. 2, 2019; accepted Aug. 6, 2019.
Author contributions: T.R. performed research; T.R., J.S.C., and A.N. analyzed data; T.R. wrote the first draft of the paper; T.R., J.S.C., and A.N. edited the paper; T.R., J.S.C., and A.N. wrote the paper; J.S.C. and A.N. designed research.
This work was supported by the Natural Sciences and Engineering Research Council of Canada to A.N. and J.S.C.
The authors declare no competing financial interests.
*J.S.C. and A.N. contributed equally to this work.
Correspondence should be addressed to Adrian Nestor at anestor@utsc.utoronto.ca.
https://doi.org/10.1523/JNEUROSCI.0471-19.2019

related potential (ERP) components sensitive to individual face processing can be modulated by the perception of multiple faces. Specifically, the amplitude of the N170 component was found to increase in response to viewing multiple faces compared with a single face and, hence, that common neural markers can inform both single and ensemble face processing.

Thus, single-face and ensemble face perception may exhibit different neural profiles, yet the nature, scope, and significance of such differences remain to be investigated. Of note in this respect, it is unclear whether and how specific neural signals reflect the visual content of summary representations associated with the perception of face ensembles. To address these issues, we appeal to pattern analyses and image reconstruction methodology as applied to EEG data.

Recent studies of individual face perception have revealed the complex temporal dynamics of facial identity processing (Ghuman et al., 2014; Vida et al., 2017; Nemrodov et al., 2019) as well as the pictorial content of face representations (Nemrodov et al., 2018). Specifically, such work has succeeded in decoding neural signals associated with different facial identities, in characterizing their temporal profile, and in reconstructing the visual appearance of single faces as perceived by different participants from corresponding EEG signals. Here, we rely on such methodology to examine the dynamics and the informational content of ensemble identity perception.

Specifically, we use EEG and behavioral data to address a series of related questions regarding: (1) the effectiveness of forming summary ensemble representations of facial identity explicitly and implicitly (i.e., with or without specific instructions); (2) the possibility of identifying and visualizing neural representations of ensemble summaries; and (3) the characterization of the temporal profile associated with ensemble perception. Briefly, our investigation supports the development of summary representations both explicitly and implicitly while pointing to the benefit and the distinctiveness of implicit processing. To our knowledge, this is the first demonstration regarding the existence of neural representations of face ensemble summaries as demonstrated by their successful decoding and reconstruction from neural data. Equally important, the present work sheds light on the dynamics of ensemble processing and its relationship with single-face processing. Thus, collectively, the present results speak to the nature, fine-grained visual content, and emergence of summary representations from visual ensembles.

Materials and Methods

Participants

Fourteen healthy adults (10 females; age range: 19–25 years) were recruited from the University of Toronto community to participate in this study. All participants were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological or visual impairment. Participants provided signed, informed consent and were financially compensated for participating in this study. This study was approved by the Research Ethics Board at the University of Toronto.

Stimuli

Stimulus selection and processing proceeded in three steps, as follows. First, color images of Caucasian males were selected from the Radboud database (Langner et al., 2010) to display young adult faces with neutral expressions and frontal pose, gaze, and illumination. These images were scaled uniformly and aligned with roughly the same position of the eyes and the nose, cropped to show only internal features of the face, and normalized with the same mean and root-mean-square contrast values for each channel in CIEL*a*b*color space.

Next, in a second step, of 60 such images, 4 groups of 6 faces were selected, so that two sets (i.e., 1–1 and 1–2) yielded a similar average face and the other two (i.e., 2–1 and 2–2) yielded a distinct similar average face. Specifically, we randomly sampled 4 groups of faces for a total of 1000 iterations under the constraint that the L2 image distance between the pixelwise averages of 6 faces from matching sets (i.e., the average of set 1–1 vs that of 1–2 and the average of set 2–1 vs that of 2–2) should be minimized. This procedure was used to determine which subsets of faces, of all images, yielded the most desirable result in terms of pixel-based, average-face distances.

Third, the average of each set was subtracted from all images in that set and replaced with the average of all 12 faces from two matching sets (e.g., the average of set 1–1 was subtracted from all individual faces in 1–1 and the average of all faces in sets 1–1 and 1–2 was added to those individual faces). Thus, our procedure aimed to deliver two sets of faces (i.e., 1–1 and 1–2) that shared the same pixelwise average face, whereas the other two (i.e., 2–1 and 2–2) shared a different average. To be clear, we note that Step 2 above already approximates a solution and, thus, limits the subsequent impact of Step 3 on image processing; for instance, if the average of selected set 1–1 were replaced by a widely dissimilar average of sets 1–1 and 1–2, this would yield image artifacts and an unrealistic appearance for the individual faces in set 1–1. In summary, the procedure above delivered a total of 26 single-face stimuli consisting of 24 unique facial identities, divided across 2 pairs of matching face sets, along with 2 corresponding average faces (see Fig. 1).

Next, for the purpose of constructing ensemble stimuli, the 6 faces from each set were displayed in a circular arrangement leading to four base face ensembles (see Fig. 1). Further, to create multiple versions of each ensemble, each display was "rotated" by shifting individual faces clockwise by one position six times. This procedure delivered a total of 24 ensemble stimuli.

In addition, for the purpose of the EEG experiment, 6 Caucasian female face images were also extracted from the same database. Single-face stimuli were generated from these images in the manner described above for male face stimuli. Further, one base ensemble, allowing 6 rotation-based versions, was constructed from these images.

Experimental design and statistical analysis

Behavioral experiment. Participants performed a one-back identity task first with single-face stimuli, for three consecutive blocks, and then with ensemble stimuli, for the next four consecutive blocks.

During the first block, each trial had the following structure: a fixation cross was displayed for 400 ms, followed by a single-face stimulus for 600 ms, then by a fixation cross for 600 ms, and by a second face stimulus for 600 ms. The second stimulus was replaced by a fixation cross which stayed on screen until the participant made a response (i.e., by pressing designated keys for "same"/"different" identities). Each single face subtended an angle of 3° × 2° at a distance of 80 cm from the screen while the head of the participant was stabilized with the aid of a chin rest. All stimuli were presented at least twice (at most 3 times) within the block in pseudorandom order, and trials with repeated versus different stimuli occurred equally often. The first block consisted of 50 trials and took ~5 min to complete. The second and third blocks were similar to the first, except that face stimuli were presented for 300 ms and each block consisted of 75 trials.

Next, participants performed an ensemble recognition task analogous to the single-face task. Specifically, participants were presented with face ensembles and were asked to determine whether two consecutive ensembles had the same or different average identities. For the purpose of this task, participants were asked to fixate a central cross at all times and to avoid looking at specific faces in an ensemble. On any trial, an ensemble from a given group could only be followed by an ensemble from a different group (e.g., if the first face ensemble was 1–2, then the following ensemble could be 1–1, 2–1, or 2–2, but not 1–2). Thus, no ensemble stimulus could be repeated within a trial and no individual face image was repeated across different ensembles in the same trial. Ensemble stimuli subtended an angle of ~9° × 7° and same-average versus different-average trials occurred equally often. During block 4 (i.e., the first ensemble block), participants viewed each stimulus for 600 ms and completed 75 trials over the course of 5 min. During blocks 5–7, stimuli were presented for only 300 ms and participants completed 75 trials per block.

All stimuli were presented against a black background on a monitor with a 1280 × 1080 resolution and a 60 Hz refresh rate. Participants completed their behavioral testing within a single 45 min session. MATLAB and Psychtoolbox 3.0 (Brainard, 1997; Pelli, 1997) (RRID: SCR_002881) were used to present stimuli, record participant responses, and analyze the data.

EEG experiment: behavioral methods. The experiment was divided across two sessions performed on separate days no more than 3 d apart. Each session, consisting of 2 training blocks and 16 experimental blocks, lasted ~3 h (including EEG equipment setup).

Each experimental block could involve either single-face presentations (i.e., single-face blocks) or ensemble presentations (i.e., ensemble blocks). In each session, the two types of blocks were interleaved in groups of four as follows: four single-face blocks were followed by four ensemble blocks, which were followed by four single-face blocks and then four ensemble blocks. Single-face blocks contained 234 trials, including 26 male identities repeated 8 times and presented in a pseudorandom order, as well as 26 catch trials consisting of female faces. Similarly, face ensemble blocks contained 216 total trials, including 24 male ensemble stimuli repeated 8 times and presented in pseudorandom order, as well as 24 catch trials consisting of female face ensembles. For both types of block each stimulus was displayed for 300 ms followed by a fixation cross with a variable 600–700 ms duration, and the task of the participants was to respond to female faces by pressing a key as soon as they were displayed. In all other respects, stimulus presentation followed the procedure described above for the behavioral experiment.

Before EEG data collection, participants also completed one training block of each type (i.e., one single-face block followed by a face-ensemble block).

EEG experiment: data acquisition and preprocessing. EEG data were recorded using an ActiveTwo EEG recording system (Biosemi). The electrodes were arranged according to the International 10/20 System, and the electrode offset was kept <40 mV. The EEG was low-pass filtered using a fifth-order sync filter with a half-power cutoff at 204.8 Hz and then digitalized at 512 Hz (i.e., 512 time bins per second, ~1.95 ms per time bin) with 24 bits of resolution. All data were digitally filtered offline (zero-phase 24 dB/octave Butterworth filter) with a bandpass of 0.1–40 Hz.

Next, data were separated into epochs from 100 ms before stimulus presentation until 900 ms after stimulus presentation. Epochs corresponding to go trials (i.e., single female faces and female face ensembles), false alarms, and misses were discarded from analysis. Further, noisy electrodes were interpolated if necessary (no more than 3 per participant), and epochs were rereferenced to the average reference. In addition, before univariate ERP analyses, data were inspected for artifacts and, using Infomax ICA (Delorme et al., 2007), ocular artifacts, such as eye-blinks, were removed. After removing trials containing artifacts and/or false alarms, we retained an average of 98% trials across participants (range: 96%–99%); we note the relatively small number of false alarms as participants performed the go/no-go recognition tasks at ceiling (range: 92%–99% accuracy).

All analyses were performed using Letswave 6 (Mouraux and Iannetti, 2008) and MATLAB.

Univariate ERP analyses. Twelve bilateral electrodes located over homolog occipitotemporal (OT) areas (left: P5, P7, P9, PO3, PO7, and O1; right: P6, P8, P10, PO4, PO8, and O2) were used in the ERP analysis. We selected these electrodes because of the robustness of classical ERP components (e.g., N170) recorded at their location and because of their ability to support pattern discrimination of facial identity (Nemrodov et al., 2018). Data corresponding to each type of stimulus (i.e., single-face and face ensemble) were averaged across electrodes separately for each participant to create a grand average waveform. For univariate analyses, we separately identified the P1, N170, P2, and N250 components and conducted pairwise two-sample *t* tests to compare their amplitudes and their latencies (i.e., onset times) between single faces and ensembles.

Pattern classification: single-face and ensemble decoding. Pattern analyses were conducted separately for each participant. Specifically, decoding estimates were computed for each participant and then averaged and compared against each other and against chance (50%) via two-tailed *t* tests.

Spatiotemporal patterns consisting of 3684 features (12 electrodes × 307 time bins during the interval 50–650 ms) were constructed through concatenation separately for each epoch (Nemrodov et al., 2018). A relatively wide 600 ms temporal interval was selected, in this respect, to contain both early and higher-level visual information relevant to single-face processing (Ghuman et al., 2014; Nemrodov et al., 2016, 2018; Vida et al., 2017) as well as to capture the extended time course of ensemble processing (Haberman et al., 2009). All epochs corresponding to the same stimulus in a given block were then averaged to increase the signal-to-noise ratio of the observations for classification purposes (Grootswagers et al., 2017). Thus, for single-face stimuli, a maximum of eight patterns and no less than six, after EEG preprocessing, were averaged to deliver a single observation. Similarly, for ensembles, all stimuli within a block corresponding to the same base ensemble, for a maximum of 48 patterns (i.e., 6 rotations × 8 repetitions) and no less than 45, were averaged into a single observation.

Next, each spatiotemporal feature was separately *z*-scored and normalized across observations within the interval 0–1. Pairwise classification of single faces and ensembles was then conducted across these observations using linear support vector machines ($c = 1$) and leave-one-block-out cross-validation. In total, 325 single-face pairs and 6 ensemble stimulus pairs were evaluated.

Further, to assess the processing of summary identity across ensembles containing different average identities, classification was performed with different training and testing pairs (e.g., a classifier would be trained to discriminate Ensembles 1–1 and 2–1, and then tested with Ensembles 1–2 and 2–2) for a total of 4 combinations; all possible training/testing pair combinations were considered for each participant. In this case, training was performed on all available observations for one pair of ensembles, and testing was performed on all observations for the other pair. Critically, we also assessed the possibility of decoding summary ensemble representations by training a classifier on data from Ensembles 1–1 and 1–2 versus 2–1 and 2–2 and testing it on data corresponding to their single average faces from single-face blocks. As a control, we also trained a classifier on data for single faces from different ensemble groups (i.e., 1–1 and 1–2 vs 2–1 and 2–2) and tested it on data from the corresponding ensemble groups.

Last, all comparisons to chance, as described above, were also conducted against permutation-based estimates (rather than a preset 50% chance level). To this end, for each participant and type of classification, we computed chance estimates by randomly shuffling classification training labels 1000 times and by deriving and averaging a corresponding number of classification estimates. Then, we assessed significance by comparing true classification results against participant-matched chance-level estimates (two-tailed paired *t* test across participants).

Pattern classification: spatiotemporal dynamics. To estimate the time course of single and ensemble face processing the analyses above were conducted again across multiple temporal windows. Specifically, classification was conducted across ~10 ms windows (i.e., 5 time bins × ~1.95 ms = ~9.75 ms) relying on 60 feature patterns (5 consecutive time bins × 12 electrodes). The analysis was performed between −100 and 700 ms, corresponding to 410 time bins, by sliding the window one bin at a time. Thus, for each type of classification, this analysis provides a fine-grained temporal estimate of discrimination.

Further, to evaluate the cross-temporal generalizability of relevant information for any given type of classification, training was performed for every window, and then testing was conducted for every possible window (Isik et al., 2014). Temporal cross-decoding yielded a 406 × 406 matrix whose diagonal corresponds to the time course estimated above (i.e., when training and testing is performed over the same temporal window). This analysis is instrumental in assessing the redundancy/complementarity of information contained within the EEG signal across different intervals.

For both types of analysis, classification accuracy across participants was compared with chance (50%) via one-sample *t* tests FDR-corrected for multiple comparisons.

Face space and face ensembles. Face space constructs were derived separately for each participant by applying multidimensional scaling (MDS) to EEG-based estimates of face similarity. Specifically, classification accuracies based on a large temporal window (i.e., 50–650 ms) yielded a 30 × 30 confusability matrix encoding the relationship between every pair of single faces and face ensembles (i.e., 24 individual identities, two average faces, and four ensembles). Next, all values were linearly scaled between 0 and 1, and metric MDS was applied to approximate the corresponding face space. The dimensionality of the space was restricted to 15 dimensions as this was sufficient to account for most variance in our data (>82% for each participant) while also minimizing the possibility of overfitting for classification analyses performed in this space.

In addition, for completeness, face space was also estimated based on image properties to visualize the objective structure of the stimulus space. Specifically, the pixelwise Euclidean distance between pairs of single-face stimuli (i.e., 24 individual identities and two average faces) was computed in CIEL*a*b*. Then, MDS was applied to the resulting confusability matrix to derive a stimulus-based face space. To be clear, this analysis does not include face ensembles as their similarity to single faces cannot be directly measured (i.e., they involve different types of visual display).

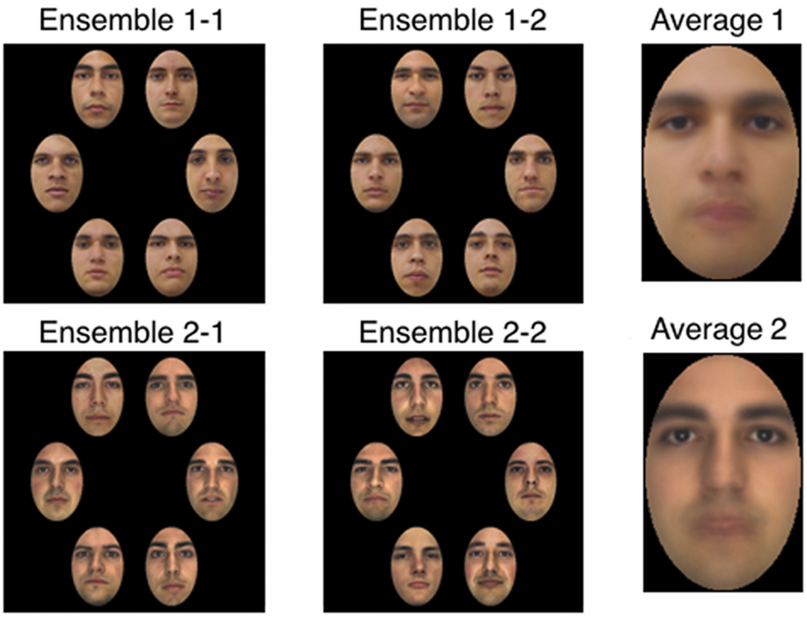

Figure 1. Experimental stimuli. A total of 24 unique faces were selected, processed, and divided into four different ensembles such that two ensembles (1–1 and 1–2) shared one average face (average 1) and the other two ensembles (2–1 and 2–2) shared a different average face (average 2).

Next, a linear discriminant analysis classifier was trained on EEG-based face space coordinates to classify single faces across the two ensemble groups (i.e., 12 faces associated with Ensembles 1–1 and 1–2 vs 12 faces from ensembles 2–1 and 2–2) and was tested on the four face ensembles. The results of this analysis are presented in Figure 6 in a 3D space for visualization purposes.

Image reconstruction of single-face and ensemble summary percepts. The procedure for facial image reconstruction relies on a recent approach capitalizing on the spatiotemporal structure of EEG data (Nemrodov et al., 2018, 2019). Here, we use this technique to assess and visualize representations of individual faces as well as, notably, representations associated with ensemble perception.

Briefly, the procedure for single-face reconstruction involves a series of steps, as follows. First, visual features accounting for the structure of face space were derived separately for each dimension of the space. These features were computed as weighted sums of image stimuli following a strategy akin to reverse correlation (Gosselin and Schyns, 2003; Murray, 2011; Smith et al., 2012). Following conversion to CIEL*a*b, face stimuli were summed proportionately to their coordinates separately for every dimension of face space. The procedure yielded a total of 15 features or classification images (CIMs), one for each corresponding dimension.

Second, for each dimension, permutation-based CIMs were generated by randomly shuffling the coefficients associated with the stimulus images. Pixel intensities in the true CIM were then compared with the corresponding intensities of pixels in the permutation-based CIMs, and only CIMs that contained pixel values significantly different from chance were retained for reconstruction purposes.

Third, the coordinates of the target face were estimated into the existing face space. Importantly, to avoid dependency, the target face was left out from face space construction and feature derivation by using a leave-one-out approach.

Last, a linear combination of significant CIMs, proportional with the coordinates of the target in face space, was added to an average face obtained from all other faces. The outcome of this procedure yields a visual approximation of the appearance of the target for a specific participant.

Ensemble summary reconstruction relied on the procedure above with two modifications. First, all individual faces were included in the approximation of face space, as there was no need to systematically leave one single face out. Second, following the same reasoning, CIMs were added to the average of all single-face stimuli.

Evaluation of image reconstructions. For single faces, reconstruction accuracies were estimated as the relative number of instances for which the reconstructed image in CIEL*a*b was more similar to its target than to any other stimulus. Reconstruction accuracy averaged across participants was then compared with chance (50%) via two-tailed one-sample t tests.

For face ensembles, reconstruction accuracy was estimated in the same manner, except that reconstructions were compared with average faces instead of individual faces. For instance, reconstructions for Ensembles 1–1 and 1–2 were labeled as accurate if they were closer to average identity 1 than to average identity 2.

Results

Experiment 1: behavioral experiment

Behavioral testing was conducted to confirm and to assess explicit sensitivity to summary face representations. To this end, 24 distinct face images were selected, processed, and divided into four ensembles of 6 faces such that two ensembles (i.e., 1–1 and 1–2) shared the same pixelwise average face and the other two ensembles (i.e., 2–1 and 2–2) shared a different average (Fig. 1). In response to such stimuli, participants were instructed to maintain central fixation and to perform a one-back identity task by indicating whether two ensembles, sequentially presented, shared the same average identity or not. In addition, participants performed a one-back image task with pairs of centrally presented single-face stimuli.

An evaluation of accuracy showed that, in the single-face task, participants reached, as expected, ceiling performance (mean score = 96.51%, range: 92%–100%, SD = 2.67%; two-tailed one-sample t test against 50% chance: $t_{(13)} = 65.18$, $p < 0.0001$, Cohen's $d = 17.42$). In the ensemble task, despite the considerable level of difficulty reported by participants, performance (mean score = 60.79%, range: 54%–67.7%, SD = 1.06%) remained above chance (two-tailed one-sample t test; $t_{(13)} = 10.16$, $p < 0.0001$, $d = 2.71$) but was not as accurate as performance in the single-face task (two-tailed paired *t* test; $t_{(13)} = -40.94$, $p < 0.0001$, $d = 10.55$). However, the comparison of the two tasks showed that accuracy was correlated across participants (Pearson correlation; $r = 0.577$, $p = 0.03$).

Consistent with the results above, an examination of reaction times (RT) found that, relative to the single-face task (mean RT = 409 ms; range: 291–614 ms), the ensemble task (mean RT = 827 ms, range: 550–1298 ms) yielded significantly longer RTs (two-tailed Wilcoxon signed-rank test; $z = 3.30$, $p < 0.01$, $r = 0.88$). No significant correlation was found for RTs between the two tasks across participants ($r = 0.60$, $p = 0.84$).

In agreement with previous work, the current results indicate that participants are capable of extracting a summary representation of facial identity from ensemble stimuli. However, in contrast to prior work, participants did not directly match an ensemble to a summary representation but, rather, to a different ensemble consisting of a different group of faces. Hence, we provide a novel demonstration of summary identity encoding, and we show that summary representations are robust enough to be reliably compared, even across different ensembles.

Experiment 2: EEG experiment

To assess visual representations of summary identity and their temporal dynamics, EEG data were collected with the same group of participants. Specifically, EEG data were recorded while participants viewed sequences of single faces and face ensembles presented for 300 ms each. Participants were instructed to maintain central fixation and to perform a go/no-go task by pressing a key whenever they saw a single female face in single-face blocks, or an ensemble of female faces in ensemble blocks.

Univariate analyses

Multiple ERP components (P1, N170, P2, and N250), typically related to face processing, were identified across 12 bilateral OT electrodes (left: P5, P7, P9, PO3, PO7, and O1; right P6, P8, P10, PO4, PO8, and O2); these electrodes were selected based on their known relevance for face processing (e.g., robust N170 components) (Itier and Taylor, 2002; Caharel et al., 2009; Ince et al., 2016) and ability to support face decoding (Nemrodov et al., 2016). These components were compared across single faces and face ensembles to assess coarse differences in the ERP signal and to relate the current results with prior ERP investigations (Puce et al., 2013). Specifically, differences in amplitude and latency were evaluated across single faces and ensembles for each component (two-tailed paired *t* tests). These analyses revealed lower P1 amplitude ($t_{(13)} = -2.40$, $p = 0.032$, $d = 0.218$), earlier N170 ($t_{(13)} = -5.63$, $p < 0.0001$, $d = 0.400$), and earlier P2 ($t_{(13)} = -3.00$, $p = 0.011$, $d = 0.283$) components for face ensembles relative to single faces (Fig. 2). No other comparisons reached significance (all *p* values > 0.10).

To relate behavioral findings with the present results, performance accuracy in the ensemble task was correlated across participants with P1 amplitude as well as with the onset of N170 and P2. However, no correlations reached significance (P1: $r = -0.47$, $p = 0.87$; N170: $r = 0.43$, $p = 0.13$; and P2: $r = -0.34$, $p = 0.23$).

Single-face decoding

Pairwise face classification was conducted across spatiotemporal patterns (i.e., 12 OT electrodes, 50–650 ms after stimulus onset) to estimate the overall discriminability of different stimuli and of their underlying representations. First, this analysis was conducted across each pair of individual faces (i.e., 276 pairs corresponding to 26 faces). The analysis yielded above-chance classification (mean accuracy = 63.50%, SD = 5.84%; one-sample *t* test against 50% chance; $t_{(13)} = 8.64$, $p < 0.0001$, $d = 2.31$), consistent with the ability of the EEG signal to capture identity-related facial information (Nemrodov et al., 2016).

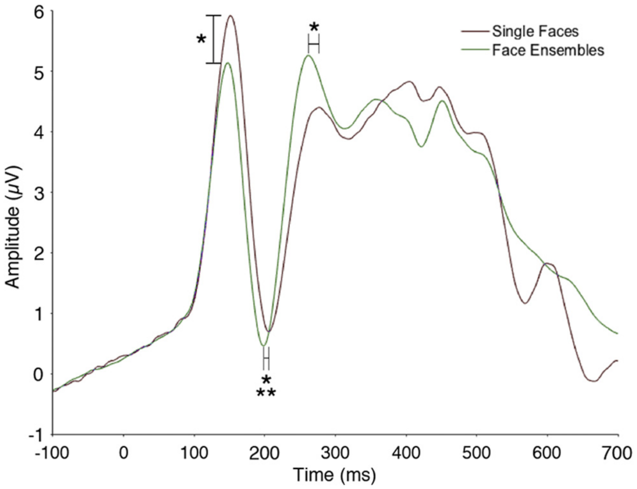

Figure 2. ERP waveforms for single faces and face ensembles. A comparison of ERPs elicited in response to single faces and face ensembles across 12 bilateral occipitotemporal electrodes revealed a reduced amplitude for the P1 component, an earlier N170 component, and an earlier P2 component for face ensembles relative to single faces *$p < 0.05$. ***$p < 0.001$.

Second, we considered the impact of our stimulus design procedure, which likely amplified visual differences between faces purposed for the construction of different ensemble groups (i.e., the visual similarity of faces from Ensembles 1–1 and 1–2 compared with faces from 2–1 and 2–2) (Fig. 1). Hence, it is possible that successful face decoding was driven exclusively by pairs of faces from different groups. To evaluate this possibility, single-face classification was evaluated for all possible face pairs within each group (i.e., 132 pairs) and separately for all possible face pairs across groups (i.e., 144 pairs). As expected, classification accuracy was higher for the latter compared with the former (two-tailed paired *t* test; $t_{(13)} = 6.92$, $p < 0.0001$, $d = 1.03$). However, classification accuracy was significantly above chance both across groups (mean accuracy: 63.50%, SD = 5.85%, one-sample *t* test against chance; $t_{(13)} = 8.64$, $p < 0.0001$, $d = 2.31$) and within groups (54.30%, SD = 4.15, $t_{(13)} = 3.87$, $p = 0.0019$, $d = 1.03$) (Fig. 3*A*). Thus, successful decoding was not driven exclusively by pairs of faces from different groups. Importantly, within-group faces can be discriminated from each other, and this should not prevent, in itself, the discrimination of same-group ensembles (e.g., Ensembles 1–1 and 1–2).

Last, we note that comparisons of decoding accuracy against permutation-based chance levels replicated qualitatively all classification results reported above.

Ensemble decoding

Ensemble classification was assessed across all 6 possible ensemble pairs. Classification accuracy was above chance (mean classification = 58.07%, SD = 4.17; one-sample *t* test; $t_{(13)} = 7.23$, $p < 0.0001$, $d = 1.93$), indicating overall sensitivity to visual ensemble information.

However, while within-group single faces are discriminable, as seen above, it is possible that same-group ensembles are not. Specifically, if face ensembles are summarized into a single identity representation, as suggested by prior behavioral work (de

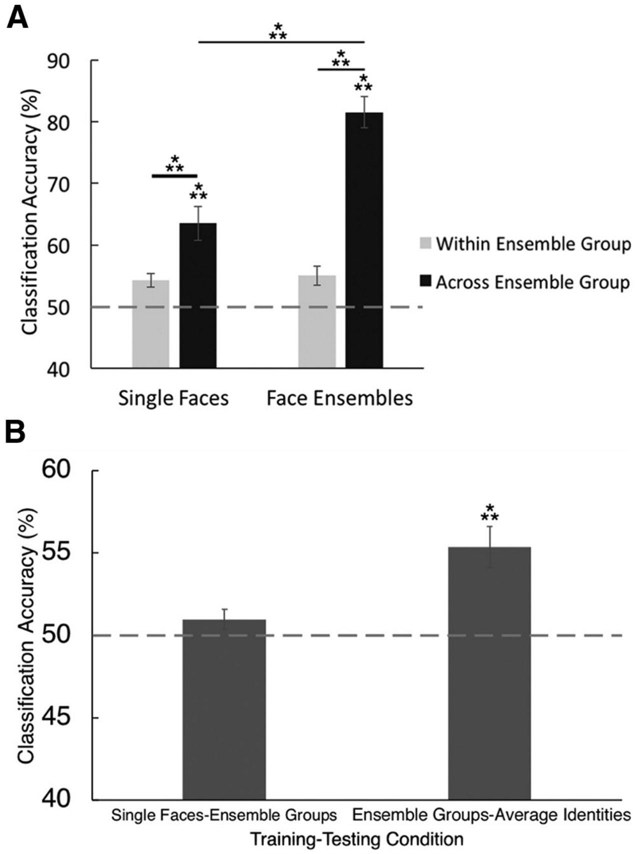

Figure 3. EEG-based decoding results. ***A***, Classification accuracies for single faces and face ensembles were estimated and compared both within group (i.e., same average identity) and across groups (i.e., different average identities). Ensemble decoding yielded higher classification accuracy than single-face decoding, and cross-group decoding yielded higher accuracy than within-group decoding. ***B***, Cross-decoding accuracies were estimated by (left) training a classifier on pairs of single faces from different groups (i.e., 1–1 and 1–2 vs 2–1 and 2–2) and testing on ensembles from the corresponding groups as well as by (right) training on ensembles from different groups and then testing on average faces. Only the latter analysis yielded significant decoding accuracy. Error bars indicate ± 1 SE across participants. $^{***}p < 0.001$.

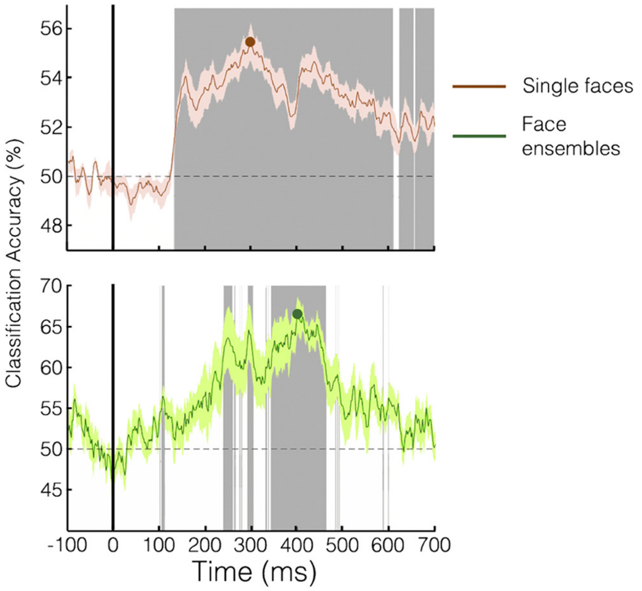

Figure 4. The time course of EEG-based classification accuracy. Classification was computed across consecutive 10 ms windows >12 occipitotemporal electrodes and averaged across all participants. Both single faces (top) and ensembles (bottom) exhibited extended intervals of above-chance accuracy. The time course of single-face classification achieved above-chance accuracy at 138 ms, peaked at 310 ms, and maintained significance until 605 ms ($p < 0.05$; FDR correction across time, $q < 0.05$). The time course of face-ensemble classification achieved above-chance classification accuracy at 101 ms, exhibited several shorter windows of significance across the time course ($p < 0.05$; FDR correction across time, $q < 0.05$), and peaked at 408 ms. Shaded areas represent ± 1 SE across participants. Circles represent overall peaks of classification accuracy for single faces and face ensembles.

Fockert and Wolfenstein, 2009; Neumann et al., 2013, 2017; Haberman et al., 2015), same-group ensembles (e.g., Ensembles 1–1 and 1–2) may not be discriminable from each other if, though consisting of different individual faces, they are perceptually reduced to the same average identity. To evaluate this possibility, ensemble classification was evaluated separately for same-group ensembles (i.e., 2 pairs) and cross-group ensembles (i.e., 4 pairs) (Fig. 3*A*). Interestingly, same-group ensemble classification was only marginally significant (mean accuracy = 55.02%; two-tailed one-sample *t* test against 50%, $t_{(13)} = 1.85$, $p = 0.060$, $d = 0.50$), in contrast to cross-group ensembles (mean accuracy = 81.53%; $t_{(13)} = 12.50$, $p < 0.0001$, $d = 3.34$); these results were also confirmed by comparisons against permutation-based chance, which yielded qualitatively similar results (i.e., significantly above-chance decoding for cross-group ensembles but not for same-group ensembles).

Further, to relate single and ensemble face decoding, we compared their levels of decoding accuracy. A two-way repeated-measures ANOVA (2 stimulus types: single face vs ensemble, and 2 classification groups: within and across group) revealed significant main effects of stimulus type ($F = 65.60$, $p < 0.0001$, $\eta^2 = 0.84$), with higher accuracy for ensembles relative to single faces, and classification group ($F = 125.69$, $p < 0.0001$, $\eta^2 = 0.91$), with higher accuracy for cross-group versus within-group decoding, as well as a significant interaction between the two factors ($F = 383.05$, $p < 0.0001$, $\eta^2 = 0.97$). A further comparison of cross-group decoding accuracy with the two types of stimuli revealed no significant correlation across participants ($r = 0.17$, $p = 0.56$).

In addition, we computed correlations between behavioral accuracy in the ensemble task with both overall ensemble decoding accuracy and cross-group ensemble decoding accuracy. However, neither correlation reached significance ($r = 0.020$, $p = 0.95$ and $r = -0.25$, $p = 0.39$, respectively).

Summary representation decoding: cross-stimulus classification of face ensembles

As a more robust test of ensemble representations, we assessed the ability to decode the same summary representation from different ensembles. To this end, we trained the classifier on every possible combination of two ensembles with distinct average identities (e.g., training on Ensembles 1–1 and 2–1), and tested it on the remaining two ensembles, which matched the former with respect to corresponding average identities (e.g., testing on Ensembles 1–2 and 2–2). This analysis yielded 81.53% mean accuracy (SD = 9.44%; one-sample *t* test against chance, $t_{(13)} = 12.50$, $p < 0.0001$, $d = 3.34$). These results are convergent with the outcome of ensemble decoding relying on the same ensembles for training and testing purposes, as described above (Fig. 3*A*), and provide, by means of cross-stimulus classification, a more stringent test of sensitivity to summary representations.

Further, we investigated whether the patterns corresponding to a single average face could be correctly classified based on the patterns from its corresponding ensembles. To this end, we

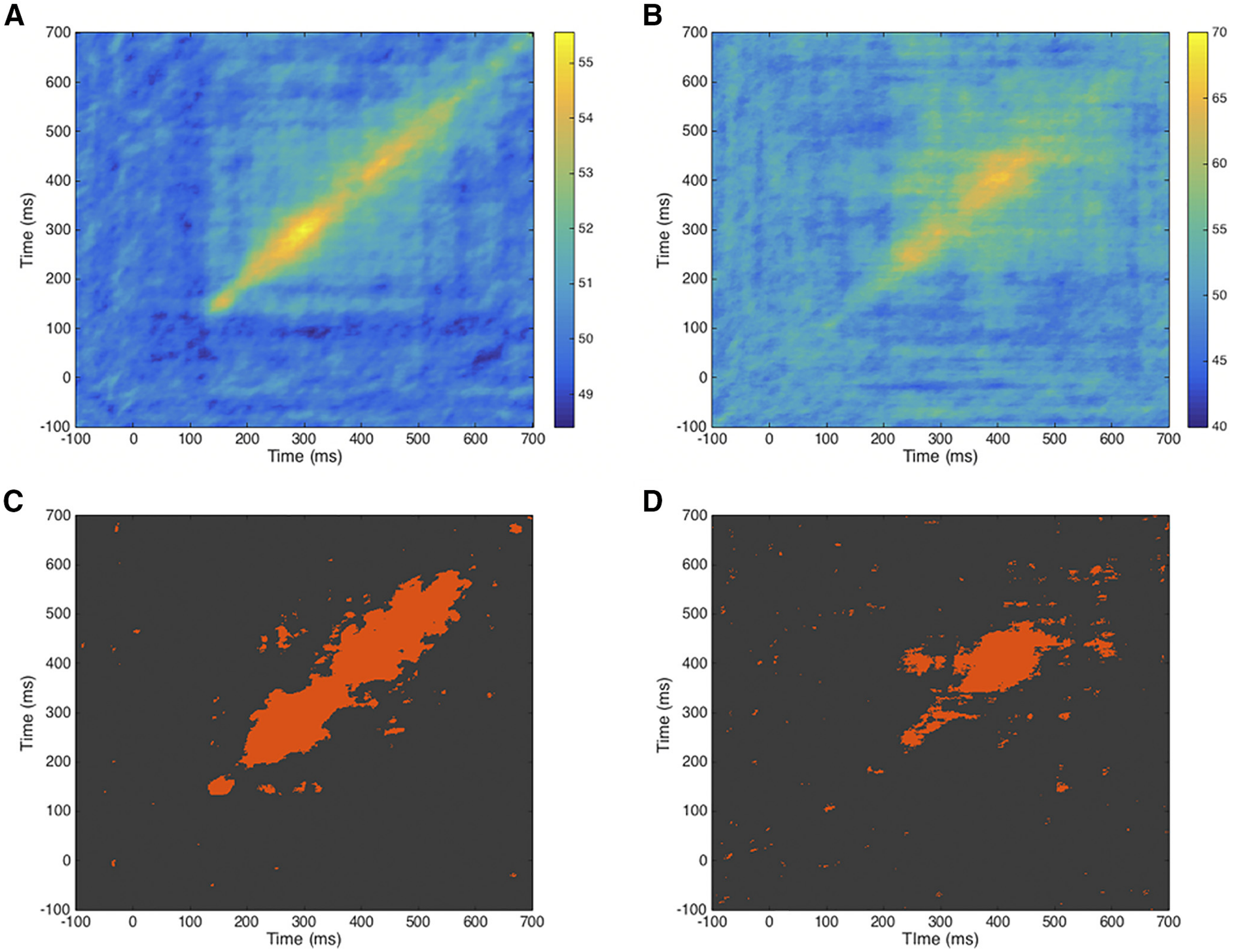

Figure 5. Cross-temporal generalizability of single-face and ensemble processing. Cross-temporal decoding for (***A***) single faces and (***B***) face ensembles were estimated by training on 10 ms intervals (*x* axis) and testing on any 10 ms interval (*y* axis). ***C***, ***D***, Time points that are significantly above chance (marked with red) for single faces and face ensembles, respectively (comparison to 50%, FDR correction, $q < 0.05$). Limited generalizability is found by above-chance decoding in the proximity of training intervals between (***C***) 150 – 600 ms for single faces and (***D***) 225–550 ms for ensembles.

trained a pattern classifier on signals corresponding to the two ensemble groups (i.e., Ensembles 1–1 and 1–2 vs 2–1 and 2–2) and tested it on the signals from their corresponding single average identities (i.e., average face 1 vs average face 2) (Fig. 1). Classification accuracy was above chance (55.35%; SD = 4.66%, $t_{(13)} = 4.31$, $p < 0.0001$, $d = 1.15$) (Fig. 3*B*).

While the results above support the neural encoding of summary representations, an alternative account relies on the encoding of individual faces. Specifically, it is possible that individual faces from different ensemble groups have sufficient within-group similarity and between-group dissimilarity that the encoding of any two individual faces from different groups may suffice for the purpose of decoding the corresponding ensembles. To assess this possibility, we attempted to classify face ensembles by training a classifier on the data from any two single identities belonging to different ensemble groups, and then tested the classifier on the two ensemble groups. This analysis yielded chance-level performance (mean classification accuracy = 50.97%, SD = 2.20%; one-sample *t* test, $t_{(13)} = 1.66$, $p = 0.12$, $d = 0.44$).

Again, to verify the validity of our results above, we compared cross-decoding accuracy against permutation-based chance levels. These additional analyses replicated qualitatively all classification results reported above.

Thus, the use of cross-stimulus classification provides evidence for the neural encoding of summary face representations. Of note, these results cannot be explained away by the encoding of unique pairs of single-face constituents within such ensembles.

Temporal dynamics of single and ensemble face processing

To elucidate the time course of face processing, classification was performed across 10 ms windows separately for single faces and face ensembles. For single faces, classification first reached significance at 138 ms, peaked at 310 ms, and exhibited an extended interval of above-chance performance (one sample *t* tests against chance, FDR-corrected across time bins, $q < 0.05$; $p = 0.031$) (Fig. 4). In contrast, for face ensembles, classification reached significance earlier, at 101 ms, it peaked later, at 408 ms, and it exhibited multiple shorter intervals of above-chance accuracy, with the longest between 360 and 460 ms ($q < 0.05$; $p = 0.01$) (Fig. 4). Also, we note that face ensembles exhibited a more gradual increase in accuracy over time compared with single faces consistent with a process of information accumulation (Haberman et al., 2015).

To examine whether complementary information exists at different time points, we compared decoding participant-specific estimates at the time of group-based peaks (i.e., at 310 ms for single faces and 408 ms for ensembles) with the corresponding results from temporally cumulative analyses (i.e., across 50–650 ms). This comparison revealed that cumulative decoding surpasses its temporal counterpart. Concretely, cumulative analysis provided an advantage both for single faces (two-tailed *t* test across participants, $t_{(13)} = 6.47$, $p < 0.001$, $d = 1.89$) and for ensembles ($t_{(13)} = 4.45$, $p < 0.001$, $d = 1.91$) consistent with the hypothesis of complementary information becoming available over time.

As a further test of this hypothesis, we examined cross-temporal generalizability, by conducting pattern classification for every possible combination of temporal windows for training and testing purposes. Figure 5 summarizes these results separately for single faces and face ensembles. Upon inspection, it appears that some degree of generalization is present, especially between 150 and 350 ms for single faces (Fig. 5*A*), and between 350 and 450 ms for face ensembles (Fig. 5*B*). However, after correcting for multiple comparisons ($q < 0.05$), generalizability appears rather limited around the diagonal both for single faces (Fig. 5*C*) and for ensembles (Fig. 5*D*).

Together, these results demonstrate that single faces and face ensembles exhibit different temporal dynamics, but their processing is similar in that largely complementary information across time appears to support successful classification for either class of stimuli.

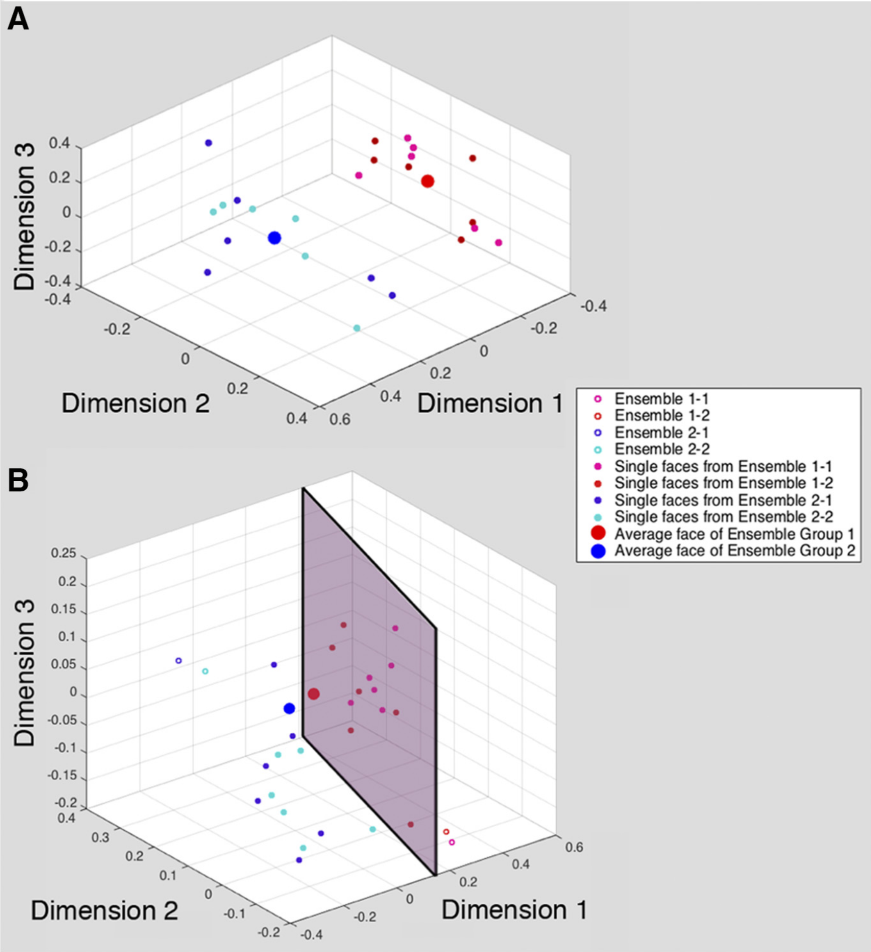

Figure 6. Face space. An approximation of face space was derived from (***A***) image-based pixelwise distances across face images and (***B***) EEG-based dissimilarity estimates associated with face perception averaged across participants (only 3 dimensions are shown for visualization purposes). Both spaces contain single faces, including average faces, whereas neural-based face space also includes representations of the four ensembles. In both spaces, single faces are segregated consistent with their membership to different groups (i.e., Ensembles 1–1 and 1–2 vs Ensembles 2–1 and 2–2). A classifier trained on single-face representations from different groups correctly discriminates between ensemble representations by placing them on the corresponding side of the classification hyperplane (purple shading).

Face space: single faces and face ensembles

Face space provides an informative way of evaluating and visualizing the structure of face representations by capturing the pairwise similarity of different facial identities (Valentine, 1991). Specifically, the relative distance between different points, corresponding to different faces, matches their degree of behavioral-based similarity, neural-based similarity, or image similarity (O'Toole et al., 2018). Here, we appeal to this framework to evaluate the relationship between neural-based ensemble summary representations and single-face representations.

To this end, for each participant, we constructed a confusability matrix containing EEG-based pairwise similarity estimates across individual faces and ensembles (i.e., 24 individual identities, two average faces, and four ensembles). Then, we estimated a face space construct by applying metric MDS across such matrices.

For visualization purposes, Figure 6*A* displays the results of such an analysis based on pixelwise image differences across stimuli, whereas Figure 6*B* displays the outcome of the EEG-based analysis averaged across participants. As expected, both instances show single-face clustering based on ensemble group. In addition, consistent with our decoding results, an examination of neural-based face space reveals that ensembles from the same groups are closer to each other than ensembles from different groups. More importantly, ensemble representations are closer to single-face representations belonging to their corresponding group. That is, representations associated with Ensembles 1–1 and 1–2 are closer to representations of single faces from those ensembles compared with single faces from the ensembles belonging to the other group, and likewise for Ensembles 2–1 and 2–2.

To evaluate more thoroughly the observation above, neural-based face spaces were approximated separately for each participant. Then, a linear discriminant analysis classifier was trained on face space coordinates to classify single faces across the two ensemble groups (i.e., 12 faces associated with Ensembles 1–1 and 1–2 vs 12 faces from Ensembles 2–1 and 2–2) and was tested on the four ensembles. This analysis found above-chance classification for each participant (average classification accuracy = 68.08%; SD = 6.62%; one-sample *t* test against chance; $t_{(13)} = 10.22$, $p < 0.0001$, $d = 2.73$). For illustration purposes, an exam-

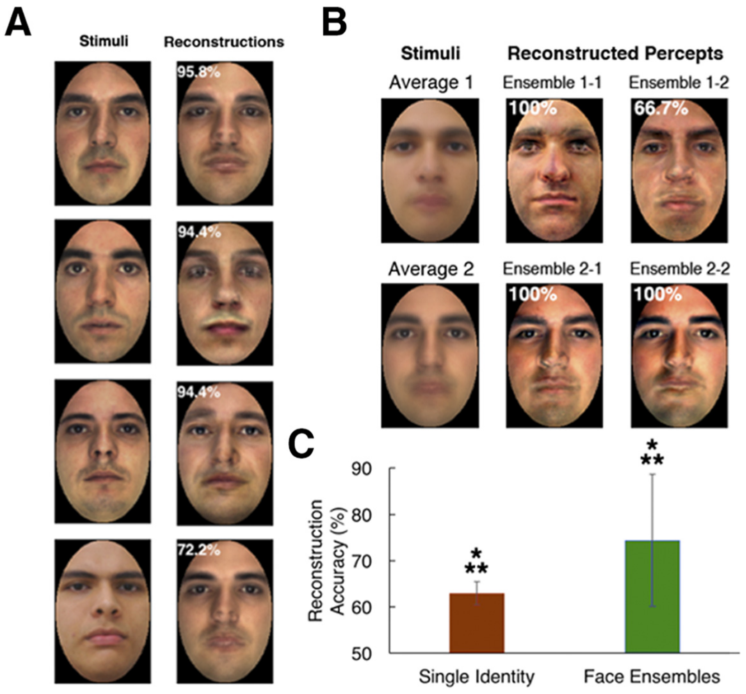

Figure 7. Neural-based image reconstruction. Examples of reconstructions and corresponding stimuli from a representative participant for (***A***) single faces and (***B***) ensemble summary representations. Reconstructions were based on EEG signals corresponding to ensemble perception but evaluated relative to ensemble average images. Numbers in the top left corner of each reconstructed image indicate image-based reconstruction accuracy. ***C***, Average reconstruction accuracies across participants. Error bars indicate ± 1 SE. ***$p < 0.001$. Additional examples of reconstructions can be found in Figure 7-1 (available at https://doi.org/10.1523/JNEUROSCI.0471-19.2019.f7-1).

ple of the linear discriminant analysis-based hyperplane is shown in Figure 6*B* for neural confusability data averaged across all participants. In this case, the hyperplane that divides face space between the two groups of single faces classifies ensemble representations with perfect accuracy.

Image reconstruction of single-face and ensemble summary percepts

Recent advances in EEG-based image reconstruction have provided the opportunity to assess and visualize representations of individual faces (Nemrodov et al., 2018, 2019). Here, we apply this methodology to approximate the visual content of summary representations in addition to that of single-face representations. Specifically, we derive visual features from the structure of face space and then combine these features to reconstruct the appearance of face representations.

This approach yielded, for each participant, reconstructions of percepts associated with single faces (Fig. 7*A*) and with ensemble summary representations (Fig. 7*B*; for additional examples, see Fig. 7-1, available at https://doi.org/10.1523/JNEUROSCI.0471-19.2019.f7-1). An image-based evaluation of individual face reconstructions relative to single-face stimuli yielded above-chance accuracy (mean accuracy = 62.93%; one-sample *t* test against 50% chance, $t_{(13)} = 11.97$, $p < 0.0001$, $d = 3.20$). More importantly, the accuracy of ensemble summary percept reconstructions, evaluated relative to average faces, was also above chance (mean accuracy = 74.40%; $t_{(13)} = 3.93$, $p = 0.00086$, $d = 1.00$).

The current analysis indicates that single-face representations reflect, as expected, the visual properties of their corresponding stimuli. More importantly, it shows that ensemble representations capture the summary visual properties of the face ensembles. Thus, the present findings confirm the summary nature of neural ensemble representations and provide a novel means to visualize their content.

Discussion

The current study investigates the neural basis of face ensemble processing with focus on summary representations and temporal dynamics. Our investigation evinced several notable outcomes, as follows.

First, we find that, behaviorally, participants are able to explicitly match summary representations across different ensembles. These results are consistent with previous work (de Fockert and Wolfenstein, 2009; Neumann et al., 2013, 2017; Haberman et al., 2015), which has demonstrated sensitivity to summary visual properties of face ensembles by the successful matching of ensembles to average face stimuli. However, it is possible that the sensitivity demonstrated in such studies is facilitated, or even developed, due to the availability of average faces, in their role of matching stimuli, throughout experimental testing. Here, we address this concern, and we provide a more stringent test of such sensitivity by showing that summary representations are robust enough to support direct ensemble comparison.

Second, we show that EEG patterns can be used to decode not only single faces (Nemrodov et al., 2018, 2019), but also face ensembles. Interestingly, decoding was successful across ensembles with different average identities, but not across ensembles with the same average face and different face constituents. This finding is consistent with the hypothesis that ensembles are largely reduced to summary representations (Haberman et al., 2015). Specifically, while information about individual constituents may not be entirely discarded (Neumann et al., 2017), summary percepts provide a convenient way of representing a wealth of information, especially when such information is only briefly available (e.g., 300 ms in our experiment). Thus, the neural signatures of individual constituent faces may be missing or considerably diminished and, hence, unable to support the decoding of distinct ensembles with the same summary representation.

Third, cross-decoding across ensembles and average face stimuli provides further evidence for summary representations. Specifically, we find that a classifier trained on ensembles with different average identities can successfully decode the corresponding average faces from each ensemble. This result is significant in that ensemble faces and single faces are presented at different positions in the visual field (i.e., parafoveally vs centrally). As eccentricity is a well-known principle of cortical organization, the two types of stimuli are likely to recruit initially different areas of early and high-level visual cortex (Hasson et al., 2002). Successful cross-decoding indicates that position-invariant neural representations of face summaries can be extracted from ensemble displays and rendered comparable to representations of single faces. Thus, decoding appears to exploit, at least to some extent, visual properties of higher-level position-invariant face representations available at later stages within the visual processing hierarchy.

Fourth, while face space (Valentine, 1991; O'Toole et al., 2018) provides a classical framework in the study of face perception, its use in the research of face ensembles has not been explored yet. Here, we embed summary ensemble representations into face space constructs both for visualization purposes and as an intermediary step in the image reconstruction procedure. An examination of face space topography is consistent with our decoding results by pointing to the separability of ensemble representations in this space. Specifically, we find that ensemble representations are positioned alongside the representations of individual faces belonging to such ensembles. Further, ensembles sharing the same average are relatively close in face space convergent with the difficulty of their decoding.

Fifth, we visualize and assess mental constructs associated with summary representations by reconstructing their visual appearance from neural data elicited by ensemble perception. Of note, this marks a clear departure from previous uses of image reconstruction aimed at retrieving the appearance of specific stimuli, whether alphanumeric characters (Thirion et al., 2006; Miyawaki et al., 2008), scenes (Naselaris et al., 2009; Nishimoto et al., 2011), or faces (Lee and Kuhl, 2016; Chang and Tsao, 2017; Zhan et al., 2019). In contrast, here we retrieve the visual content of internal representations derived from the structure of ensemble displays as opposed to that of the ensembles themselves. Our results confirm that summary representations reliably capture aspects of ensemble averages. Thus, our work provides insights into the fine-grained pictorial content of summary representations; and further, it paves the way for future work to explore exactly what and how different visual cues (e.g., shape and surface properties) are integrated into such representations.

Next, with regard to temporal dynamics, our univariate analyses confirmed the sensitivity of traditional ERP components, such as N170 (Puce et al., 2013) as well as P1 and P2, to ensemble processing. However, we note that pattern analyses of neural data afford a more thorough and robust assessment of temporal profiles (Isik et al., 2014; Cichy et al., 2014). Accordingly, our decoding results point to widely different neural profiles for single faces versus face ensembles. Specifically, for single faces, we note a first peak of decoding accuracy at 163 ms, in the proximity of the N170 component, and an overall peak at 310 ms. In contrast, ensemble processing exhibits a more gradual increase in the decoding accuracy with a peak occurring at 408 ms after stimulus onset. Limited temporal generalization also suggests largely different information supporting decoding at different time points, although a cloud of above-chance accuracy centered ~400 ms along with smaller subsequent patches of significant decoding were also noted (Fig. 5). Interestingly, the extensive time course of ensemble processing revealed here is broadly consistent with the temporal integration of information during ensemble face perception found by previous behavioral work, including the estimation of a time constant of ~800 ms underlying visual integration (Haberman et al., 2009). Thus, pattern analysis of EEG data has the ability to shed light on the dynamics of spatial ensemble processing and paves the way to analogous investigations into serial dependence in face perception (Fischer and Whitney, 2014; Liberman et al., 2014).

Importantly, the temporal profile described above is not at odds with prior research documenting the ability to rapidly extract summary representations from an ensemble, often in <100 ms (Haberman and Whitney, 2009; Li et al., 2016). Indeed, we also note a first peak in ensemble decoding at ~100 ms (Fig. 4, bottom). Rather, our findings suggest a gradual process of deriving face ensemble representations that are not fully developed until ~400 ms. Of note, this interpretation also converges with the derivation of a higher-level summary representation of ensemble identity at later stages of visual processing, as discussed above.

Related to the speed of accessing summary representations, a number of studies have proposed that such representations are developed automatically, without the explicit deployment of attention (Alvarez and Oliva, 2009; de Fockert and Wolfenstein, 2009). However, this conclusion has been debated as ensemble processing, at least in the context of letter stimuli, appears to require attention (Cohen et al., 2016; Jackson-Nielsen et al., 2017). As participants were not provided with any explicit instructions regarding ensemble summary identity in the EEG experiment, our findings support the idea that facial summary representations can be developed implicitly and automatically.

Interestingly, participants were also able to process summary representations in the behavioral experiment when provided with explicit instructions to match same-average ensembles. However, participants reported considerable difficulty with the task while behavioral accuracy was lower than that achieved by EEG-based decoding and, also, uncorrelated with it across participants. One explanation in this respect relies on the hypothesis that attention markedly changes how ensembles are processed (Chong and Treisman, 2005; Cant and Xu, 2015). Specifically, the deployment of attention may have an adverse impact on the efficiency and the robustness of ensemble processing. Future studies will need to evaluate this possibility in detail.

In conclusion, we provide behavioral and neural evidence for ensemble summary representations, we characterize their fine-grained visual content, and we take steps toward elucidating the temporal profile of their processing. Thus, the present findings serve to further our understanding of ensemble processing with regard to its representational basis, its underlying mechanisms, and its temporal dynamics.

References

Alvarez GA, Oliva A (2009) Spatial ensemble statistics are efficient codes that can be represented with reduced attention. Proc Natl Acad Sci U S A 106:7345–7350.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Caharel S, Jiang F, Blanz V, Rossion B (2009) Recognizing an individual face: 3D shape contributes earlier than 2D surface reflectance information. Neuroimage 47:1809–1818.

Cant JS, Xu Y (2012) Object ensemble processing in human anterior-medial ventral visual cortex. J Neurosci 32:7685–7700.

Cant JS, Xu Y (2015) The impact of density and ratio on object-ensemble representation in human anterior-medial ventral visual cortex. Cereb Cortex 25:4226–4239.

Cant JS, Xu Y (2017) The contribution of object shape and surface properties to object ensemble representation in anterior-medial ventral visual cortex. J Cogn Neurosci 29:398–412.

Chang L, Tsao DY (2017) The code for facial identity in the primate brain. Cell 169:1013–1028.e14.

Chong SC, Treisman A (2003) Representation of statistical properties. Vision Res 43:393–404.

Chong SC, Treisman A (2005) Statistical processing: computing the average size in perceptual groups. Vision Res 45:891–900.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

Cohen MA, Dennett DC, Kanwisher N (2016) What is the bandwidth of perceptual experience? Trends Cogn Sci 20:324–335.

Dakin SC, Watt RJ (1997) The computation of orientation statistics from visual texture. Vision Res 37:3181–3192.

de Fockert J, Wolfenstein C (2009) Rapid extraction of mean identity from sets of faces. Q J Exp Psychol (Hove) 62:1716–1722.

Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in

EEG data using higher-order statistics and independent component analysis. Neuroimage 34:1443–1449.

Fischer J, Whitney D (2014) Serial dependence in visual perception. Nat Neurosci 17:738–743.

Florey J, Clifford CW, Dakin S, Mareschal I (2016) Spatial limitations in averaging social cues. Sci Rep 6:32210.

Ghuman AS, Brunet NM, Li Y, Konecky RO, Pyles JA, Walls SA, Destefino V, Wang W, Richardson RM (2014) Dynamic encoding of face information in the human fusiform gyrus. Nat Commun 5:5672.

Gosselin F, Schyns PG (2003) Superstitious perceptions reveal properties of internal representations. Psychol Sci 14:505–509.

Grootswagers T, Wardle SG, Carlson TA (2017) Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. J Cogn Neurosci 29: 677–697.

Haberman J, Whitney D (2007) Rapid extraction of mean emotion and gender from sets of faces. Curr Biol 17:R751–R753.

Haberman J, Whitney D (2009) Seeing the mean: ensemble coding for sets of faces. J Exp Psychol 35:718.

Haberman J, Harp T, Whitney D (2009) Averaging facial expression over time. J Vis 9:1.1–13.

Haberman J, Brady TF, Alvarez GA (2015) Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. J Exp Psychol Gen 144:432–446.

Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34:479–490.

Im HY, Albohn DN, Steiner TG, Cushing CA, Adams RB Jr, Kveraga K (2017) Differential hemispheric and visual stream contributions to ensemble coding of crowd emotion. Nat Hum Behav 1:828–842.

Ince RA, Jaworska K, Gross J, Panzeri S, Van Rijsbergen Nicola J, Rousselet GA, Schyns PG (2016) The deceptively simple N170 reflects network information processing mechanisms involving visual feature coding and transfer across hemispheres. Cereb Cortex 26:4123–4135.

Isik L, Meyers EM, Leibo JZ, Poggio T (2014) The dynamics of invariant object recognition in the human visual system. J Neurophysiol 111: 91–102.

Itier RJ, Taylor MJ (2002) Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. Neuroimage 15:353–372.

Jackson-Nielsen M, Cohen MA, Pitts MA (2017) Perception of ensemble statistics requires attention. Conscious Cogn 48:149–160.

Langner O, Dotsch R, Bijlstra G, Wigboldus DH, Hawk ST, Van Knippenberg AD (2010) Presentation and validation of the Radboud faces database. Cogn Emot 24:1377–1388.

Lee H, Kuhl BA (2016) Reconstructing perceived and retrieved faces from activity patterns in lateral parietal cortex. J Neurosci 36:6069–6082.

Leib AY, Fischer J, Liu Y, Qiu S, Robertson L, Whitney D (2014) Ensemble crowd perception: a viewpoint-invariant mechanism to represent average crowd identity. J Vis 14:26.

Li H, Ji L, Tong K, Ren N, Chen W, Liu CH, Fu X (2016) Processing of individual items during ensemble coding of facial expressions. Front Psychol 7:1332.

Liberman A, Fischer J, Whitney D (2014) Serial dependence in the perception of faces. Curr Biol 24:2569–2574.

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. Nature 390:279–281.

Miyawaki Y, Uchida H, Yamashita O, Sato MA, Morito Y, Tanabe HC, Sadato N, Kamitani Y (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60:915–929.

Mouraux A, Iannetti GD (2008) Across-trial averaging of event-related EEG responses and beyond. Magn Reson Imaging 27:1041–1054.

Murray RF (2011) Classification images: a review. J Vis 11:2.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63:902–915.

Nemrodov D, Niemeier M, Mok JN, Nestor A (2016) The time course of individual face recognition: a pattern analysis of ERP signals. Neuroimage 132:469–476.

Nemrodov D, Niemeier M, Patel A, Nestor A (2018) The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction. Eneuro 5:ENEURO-0358.

Nemrodov D, Behrmann M, Niemeier M, Drobotenko N, Nestor A (2019) Multimodal evidence on shape and surface information in individual face processing. Neuroimage 184:813–825.

Neumann MF, Schweinberger SR, Burton AM (2013) Viewers extract mean and individual identity from sets of famous faces. Cognition 128:56–63.

Neumann MF, Ng R, Rhodes G, Palermo R (2017) Ensemble coding of face identity is not independent of the coding of individual identity. Q J Exp Psychol (Hove) 71:1357–1366.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641–1646.

O'Toole AJ, Castillo CD, Parde CJ, Hill MQ, Chellappa R (2018) Face space representations in deep convolutional neural networks. Trends Cogn Sci 22:794–809.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442.

Puce A, McNeely ME, Berrebi ME, Thompson JC, Hardee J, Brefczynski-Lewis J (2013) Multiple faces elicit augmented neural activity. Front Hum Neurosci 7:282.

Smith ML, Gosselin F, Schyns PG (2012) Measuring internal representations from behavioral and brain data. Curr Biol 22:191–196.

Sweeny T, Whitney D (2014) Perceiving crowd attention: a viewpoint-invariant mechanism to represent average crowd identity. J Vis 14:1–13.

Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, Lebihan D, Dehaene S (2006) Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage 33:1104–1116.

Valentine T (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. Q J Exp Psychol A 43:161–204.

Vida MD, Nestor A, Plaut DC, Behrmann M (2017) Spatiotemporal dynamics of similarity-based neural representations of facial identity. Proc Natl Acad Sci U S A 114:338–393.

Watamaniuk SN, Sekuler R, Williams DW (1989) Direction perception in complex dynamic displays: the integration of direction information. Vision Res 29:47–59.

Whitney D, Yamanashi Leib A (2018) Ensemble perception. Annu Rev Psychol 69:105–129.

Zhan J, Garrod OG, van Rijsbergen N, Schyns PG (2019) Modelling face memory reveals task-generalizable representations. Nat Hum Behav 3:817–826.